

OIEP

6

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/886,400

DATE: 10/10/2001
TIME: 14:07:19

Input Set : A:\DIVER1120-4 .TXT
Output Set: N:\CRF3\10102001\I886400.raw

4 <110> APPLICANT: DIVERSA CORPORATION
5 Murphy, Dennis
6 Ried, John
8 <120> TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF
USE THEREOF

11 <130> FILE REFERENCE: DIVER1120-4
13 <140> CURRENT APPLICATION NUMBER: 09/886,400
14 <141> CURRENT FILING DATE: 2001-06-20
16 <150> PRIOR APPLICATION NUMBER: 09/619,032
17 <151> PRIOR FILING DATE: 2000-07-19
19 <150> PRIOR APPLICATION NUMBER: 09/407,806
20 <151> PRIOR FILING DATE: 1999-09-20
22 <150> PRIOR APPLICATION NUMBER: 08/613,220
23 <151> PRIOR FILING DATE: 1996-03-08
25 <160> NUMBER OF SEQ ID NOS: 4
27 <170> SOFTWARE: FastSEQ for Windows Version 4.0

ENTERED

29 <210> SEQ ID NO: 1
30 <211> LENGTH: 52
31 <212> TYPE: DNA
32 <213> ORGANISM: Artificial Sequence
34 <220> FEATURE:
35 <223> OTHER INFORMATION: polynucleotide probe
37 <400> SEQUENCE: 1

52

38 ccgagaattc attaaagagg agaaattaac tatgagagcg ctcgtctttc ac
40 <210> SEQ ID NO: 2

41 <211> LENGTH: 31
42 <212> TYPE: DNA
43 <213> ORGANISM: Artificial Sequence
45 <220> FEATURE:
46 <223> OTHER INFORMATION: polynucleotide probe
48 <400> SEQUENCE: 2

31

49 cggaagatct aggttcccca ttttcacccc t
51 <210> SEQ ID NO: 3
52 <211> LENGTH: 1095
53 <212> TYPE: DNA
54 <213> ORGANISM: Thermococcus alcaliphilus
56 <220> FEATURE:
57 <221> NAME/KEY: CDS
58 <222> LOCATION: (1)...(1092)
60 <400> SEQUENCE: 3

48

61 ttg aga gcg ctc gtc ttt cac ggc aac ctc cag tat gcc gaa atc cca
62 Leu Arg Ala Leu Val Phe His Gly Asn Leu Gln Tyr Ala Glu Ile Pro
63 1 5 10 15

96

65 aag agc gaa atc cca aag gtc ata gag aag gca tac atc cca gtc atc
66 Lys Ser Glu Ile Pro Lys Val Ile Glu Lys Ala Tyr Ile Pro Val Ile
67 20 25 30

144

69 gag aca ctg att aaa gaa gaa att cct ttt ggg ctc aac ata acg ggc
70 Glu Thr Leu Ile Lys Glu Glu Ile Pro Phe Gly Leu Asn Ile Thr Gly

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/886,400

DATE: 10/10/2001

TIME: 14:07:19

Input Set : A:\DIVER1120-4 .TXT

Output Set: N:\CRF3\10102001\I886400.raw

71	35	40	45	
73	tat acc tta aag ttc ctc ccg aag gat att ata gac ctc gtt aaa ggg	192		
74	Tyr Thr Leu Lys Phe Leu Pro Lys Asp Ile Ile Asp Leu Val Lys Gly			
75	50	55	60	
77	ggc atc gcg agt gac ctg ata gag ata atc gga acg agc tac acg cac	240		
78	Gly Ile Ala Ser Asp Leu Ile Glu Ile Ile Gly Thr Ser Tyr Thr His			
79	65	70	75	80
81	gca ata ctc ccc ctc ctc ccg ctt agc aga gta gaa gca caa gtt cag	288		
82	Ala Ile Leu Pro Leu Leu Pro Leu Ser Arg Val Glu Ala Gln Val Gln			
83	85	90	95	
85	aga gat agg gaa gtt aag gaa gag ctc ttc gag ctt tct cca aag gga	336		
86	Arg Asp Arg Glu Val Lys Glu Glu Leu Phe Glu Leu Ser Pro Lys Gly			
87	100	105	110	
89	ttc tgg ctg cca gag ctc gcc tat gac ccg ata atc cct gcc ata ctg	384		
90	Phe Trp Leu Pro Glu Leu Ala Tyr Asp Pro Ile Ile Pro Ala Ile Leu			
91	115	120	125	
93	aag gac aac ggt tat gag tat cta ttc gcc gac ggg gag gcg atg ctt	432		
94	Lys Asp Asn Gly Tyr Glu Tyr Leu Phe Ala Asp Gly Glu Ala Met Leu			
95	130	135	140	
97	ttc tca gct cat ctc aac tcg gcg ata aag cca att aaa ccg ctc tat	480		
98	Phe Ser Ala His Leu Asn Ser Ala Ile Lys Pro Ile Lys Pro Leu Tyr			
99	145	150	155	160
101	cca cac ctt ata aag gcc caa agg gaa aag cgc ttt agg tac atc agc	528		
102	Pro His Leu Ile Lys Ala Gln Arg Glu Lys Arg Phe Arg Tyr Ile Ser			
103	165	170	175	
105	tat ctc ctt ggt ctc agg gag ctt agg aag gcg ata aag ctc gtt ttt	576		
106	Tyr Leu Leu Gly Leu Arg Glu Leu Arg Lys Ala Ile Lys Leu Val Phe			
107	180	185	190	
109	gaa ggt aag gta acg cta aag gca gtc aaa gac atc gaa gcc gta ccc	624		
110	Glu Gly Lys Val Thr Leu Lys Ala Val Lys Asp Ile Glu Ala Val Pro			
111	195	200	205	
113	gtt tgg gtg gcc gtg aac acg gct gta atg ctc ggc atc gga agg ctt	672		
114	Val Trp Val Ala Val Asn Thr Ala Val Met Leu Gly Ile Gly Arg Leu			
115	210	215	220	
117	cct ctt atg aat cct aag aaa gtg gcg agc tgg ata gag gac aag gac	720		
118	Pro Leu Met Asn Pro Lys Lys Val Ala Ser Trp Ile Glu Asp Lys Asp			
119	225	230	235	240
121	aac att ctt cta tac ggc acc gat ata gag ttc att ggc tat agg gac	768		
122	Asn Ile Leu Leu Tyr Gly Thr Asp Ile Glu Phe Ile Gly Tyr Arg Asp			
123	245	250	255	
125	att gca ggc tac aga atg agt gtt gag gga tta tta gag gtt ata gac	816		
126	Ile Ala Gly Tyr Arg Met Ser Val Glu Gly Leu Leu Glu Val Ile Asp			
127	260	265	270	
129	gag ctc aac tcg gaa ctg tgc ctt ccc tca gag ctg aag cac agt gga	864		
130	Glu Leu Asn Ser Glu Leu Cys Leu Pro Ser Glu Leu Lys His Ser Gly			
131	275	280	285	
133	agg gag ctc tac tta cgg act tcg agt tgg gca cca gat aag agc ttg	912		
134	Arg Glu Leu Tyr Leu Arg Thr Ser Ser Trp Ala Pro Asp Lys Ser Leu			
135	290	295	300	

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/886,400

DATE: 10/10/2001
TIME: 14:07:19

Input Set : A:\DIVER1120-4 .TXT
Output Set: N:\CRF3\10102001\I886400.raw

```

137 agg ata tgg aga gag gac gaa ggg aac gca aga ctt aat atg ctg tcc      960
138 Arg Ile Trp Arg Glu Asp Glu Gly Asn Ala Arg Leu Asn Met Leu Ser      320
139 305      310      315
141 tac aat atg agg ggc gaa ctc gcc ctt tta gcc gag aac agc gat gca      1008
142 Tyr Asn Met Arg Gly Glu Leu Ala Leu Leu Ala Glu Asn Ser Asp Ala      335
143      325      330
145 agg gga tgg gag ccc ctc cct gag agg agg ctg gat gcc ttc cgg gcg      1056
146 Arg Gly Trp Glu Pro Leu Pro Glu Arg Arg Leu Asp Ala Phe Arg Ala      350
147      340      345
149 ata tat aac gat tgg agg ggt gaa aat ggg gaa cct tag      1095
150 Ile Tyr Asn Asp Trp Arg Gly Glu Asn Gly Glu Pro
151      355      360
154 <210> SEQ ID NO: 4
155 <211> LENGTH: 364
156 <212> TYPE: PRT
157 <213> ORGANISM: Thermococcus alcaliphilus
159 <400> SEQUENCE: 4
160 Leu Arg Ala Leu Val Phe His Gly Asn Leu Gln Tyr Ala Glu Ile Pro
161 1      5      10      15
162 Lys Ser Glu Ile Pro Lys Val Ile Glu Lys Ala Tyr Ile Pro Val Ile
163      20      25      30
164 Glu Thr Leu Ile Lys Glu Glu Ile Pro Phe Gly Leu Asn Ile Thr Gly
165      35      40      45
166 Tyr Thr Leu Lys Phe Leu Pro Lys Asp Ile Ile Asp Leu Val Lys Gly
167      50      55      60
168 Gly Ile Ala Ser Asp Leu Ile Glu Ile Ile Gly Thr Ser Tyr Thr His
169 65      70      75      80
170 Ala Ile Leu Pro Leu Leu Pro Leu Ser Arg Val Glu Ala Gln Val Gln
171      85      90      95
172 Arg Asp Arg Glu Val Lys Glu Glu Leu Phe Glu Leu Ser Pro Lys Gly
173      100      105      110
174 Phe Trp Leu Pro Glu Leu Ala Tyr Asp Pro Ile Ile Pro Ala Ile Leu
175      115      120      125
176 Lys Asp Asn Gly Tyr Glu Tyr Leu Phe Ala Asp Gly Glu Ala Met Leu
177      130      135      140
178 Phe Ser Ala His Leu Asn Ser Ala Ile Lys Pro Ile Lys Pro Leu Tyr
179 145      150      155      160
180 Pro His Leu Ile Lys Ala Gln Arg Glu Lys Arg Phe Arg Tyr Ile Ser
181      165      170      175
182 Tyr Leu Leu Gly Leu Arg Glu Leu Arg Lys Ala Ile Lys Leu Val Phe
183      180      185      190
184 Glu Gly Lys Val Thr Leu Lys Ala Val Lys Asp Ile Glu Ala Val Pro
185      195      200      205
186 Val Trp Val Ala Val Asn Thr Ala Val Met Leu Gly Ile Gly Arg Leu
187      210      215      220
188 Pro Leu Met Asn Pro Lys Lys Val Ala Ser Trp Ile Glu Asp Lys Asp
189 225      230      235      240
190 Asn Ile Leu Leu Tyr Gly Thr Asp Ile Glu Phe Ile Gly Tyr Arg Asp
191      245      250      255

```

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/886,400

DATE: 10/10/2001
 TIME: 14:07:20

Input Set : A:\DIVER1120-4 .TXT
 Output Set: N:\CRF3\10102001\I886400.raw

192	Ile	Ala	Gly	Tyr	Arg	Met	Ser	Val	Glu	Gly	Leu	Leu	Glu	Val	Ile	Asp	
193									265					270			
194	Glu	Leu	Asn	Ser	Glu	Leu	Cys	Leu	Pro	Ser	Glu	Leu	Lys	His	Ser	Gly	
195			275					280					285				
196	Arg	Glu	Leu	Tyr	Leu	Arg	Thr	Ser	Ser	Trp	Ala	Pro	Asp	Lys	Ser	Leu	
197		290					295					300					
198	Arg	Ile	Trp	Arg	Glu	Asp	Glu	Gly	Asn	Ala	Arg	Leu	Asn	Met	Leu	Ser	
199	305					310					315					320	
200	Tyr	Asn	Met	Arg	Gly	Glu	Leu	Ala	Leu	Leu	Ala	Glu	Asn	Ser	Asp	Ala	
201					325					330					335		
202	Arg	Gly	Trp	Glu	Pro	Leu	Pro	Glu	Arg	Arg	Leu	Asp	Ala	Phe	Arg	Ala	
203				340					345					350			
204	Ile	Tyr	Asn	Asp	Trp	Arg	Gly	Glu	Asn	Gly	Glu	Pro					
205			355					360									

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/886,400

DATE: 10/10/2001

TIME: 14:07:21

Input Set : A:\DIVER1120-4 .TXT

Output Set: N:\CRF3\10102001\I886400.raw